Optimal Experimental Design for (Semi-)Batch Crystallization Processes

Gustavo L. Quilló,a,b Wannes Mores,a Satyajeet S. Bhonsale,a Alain Collas,b Christos Xiouras,b\* Jan F. M. Van Impea\*

aBioTeC+, Chemical and Biochemical Process Technology and Control, KU Leuven Campus Ghent, Gebroeders De Smetstraat 1, 9000 Ghent, Belgium

bChemical & Pharmaceutical Development and Supply, Janssen R&D, Turnhoutseweg

30, 2340 Beerse, Belgium

cxiouras@its.jnj.com , jan.vanimpe@kuleuven.be

Abstract

Knowledge gain in crystallization R&D is a stepwise process where conditions are systematically explored to develop a working process, and model-based approaches to facilitate tech transfer are gaining momentum. Despite being useful to reduce overall experimental effort, these techniques are often laden with statistical issues in the context of crystallization kinetics determination. This work demonstrates how to valorize a developed model through Optimal Experimental Design in a case-study of industrial semi-batch crystallization, achieving as much as 3.71 times more efficient data collection from a D-optimal experiment versus an experiment that uses linear cooling and dosing profiles while showcasing modelling nuances faced in the field.

**Keywords**: crystallization, optimal experimental design, population balance modelling.

* 1. Introduction

Crystallization from solution is present in approximately 80% of pharmaceutical development pipelines due to its capability of impurity rejection. The interplay between hydrodynamics, thermodynamics, and kinetics in (semi-)batch processing dictates the produced polymorph, process yield, and particle size distribution (PSD), direct impacting in the efficiency of filtration, washing, drying, mixing, tableting, and flow properties.

Population Balance Modeling (PBM) is the state-of-the-art mathematical framework to simulate crystallization, and it classically requires auxiliary kinetic expressions whose parameters need to be determined prior to process design and control. However, these are often ill-determined due to a combination of (1) high parameter correlation due to (nonlinear) model structure or data collinearity, causing degeneracy during regression, (2) insufficient system understanding, leading to poor model choice, (3) insufficient exploration of the feasibility domain, (4) experimental inaccessibility of some states, (5) the applied control policy. The advent of Process Analytical Technologies (PATs) mitigated issues, enabling frequent sampling of concentration through spectral techniques, and indirect assessment of the PSD through chord length distributions. Despite the capability of the setups, crystallization model development is almost always built (only) upon series of experiments with constant control policies because these designs offer more interpretability and easier troubleshooting. It is common to gather data through a factorial design, where isothermal desupersaturation decays are evaluated at various solvent composition levels, impurity content, seed loads, and stirring rates. Next, nonlinear regression problems are posed looping through each valid combination of the kinetic expressions from the model repository, out of which the best performing, physically relevant overall model defines the structure employed in all subsequent endeavors (Orosz et al., 2023). Optimal experimental design (OED) shines after this screening stage, where there is rudimentary knowledge about the system, but there is room for model improvement. Then the model candidate is used to plan experiments in a constrained experimental space reflecting the equipment capabilities and system complexity, avoiding polymorph conversions, phase-separation, excessive breakage, or agglomeration, dead growth zones, etc. (Quilló et al., 2023a). Practically, *in silico* experiments are sequentially generated and intercalated with data collection until the regressed parameters are within the confidence interval (CI) of the previous iteration.

Optimal experimental design (OED) is a procedure that maximally exploit experimental conditions based on the model sensitivity to the parameter values or to output prediction variance, depending on the optimized scalar statistical criterion. The most popular OED criteria for parameter estimation are the D-, A- and ME-criterion, which respectively embody the volume, outer hyperbox volume and conditioning number of the parameter joint confidence interval hyper ellipsoid. These (scalar) metrics are computed via Fisher Information Matrix (FIM), sigma-points, or other methods (Bhonsale et al., 2022).

When formulated for systems in differential form such as in crystallization, an open-loop control problem is created in which the discretized control policy, i.e., the temperature or dosing trajectory, provide (some of) the degrees of freedom to improve the objective function (Nagy et al., 2008; Quilló et al., 2023a). Likewise, the optimization can encompass the batch time, the initial condition and sampling time stamps for expensive or laborious offline data collection, e.g., PSD measurements. Moreover, the OED framework can be applied in the context of algebraic mixture models for solubility prediction to support kinetics investigations (Quilló et al., 2023b).

Although recent advances extended its application to continuous systems (Kilari et al., 2023), literature on OED for crystallization kinetics remains scarce and does not explore more sophisticated situations adjusting both temperature and solvent composition under activity-dependent supersaturation nor emphasize limitations or pitfalls of the technique. This work applies one iteration of the OED approach on the PBM to obtain the most informative piecewise linear control policy for a combined cooling-dosing (semi-)batch seeded crystallization through optimization of the D-criterion using the FIM approach.

* 1. Model-based Experimental Design

Let a system of ordinary differential equations (ODEs) be composed of nonlinear functions of the state vector , the controls , and parameters **,** and time . Consider there are states and parameters. Further, assume the outputs are linear functions of the states . Then, the sensitivity of the -th output with respect to the -th parameter can be retrieved with the -th row of coefficient matrices , as in Eq. (1).

|  |  |
| --- | --- |
|  | (1) |

The sensitivity matrix element is computed as in Eq. (2). The derivatives of the states with respect to the parameters at each time point are obtained by solving the augmented the original ODE system with coupled sensitivity equations.

|  |  |
| --- | --- |
|  | (2) |

The FIM approach is rooted on the local linearization of the model by truncating a Taylor series expansion around nominal parameter estimates and its link with the variance-covariance matrix of the outputs through the Cramér-Rao bound (Bhonsale et al., 2022). Assuming additive white measurement noise and outputs each sampled at regular time intervals, the is given by Eq. (3). The model output Jacobian is represented by Eq. (4). Note the matrix can be omitted for optimization purposes.

|  |  |
| --- | --- |
|  | (3) |
|  | (4) |

* 1. Population Balance Modelling via Standard Method of Moments

The 1-D population balance for a cooling-dosing crystallization in a well-mixed semi-batch reactor under negligible breakage and agglomeration is given by Eq. (5), where is the number density [#/(kg-neat solvent.m)], is time, is the growth rate [m/s], is the characteristic particle size and is the solvent mixture mass [kg-neat solvent].

|  |  |  |
| --- | --- | --- |
|  | where  | (5) |

The method of moments (SMOM) uses the moment transformation to obtain a set of ODEs from the original equation. The -th moment of is given by Eq. (6).

|  |  |
| --- | --- |
|  | (6) |

The resulting ODEs are given by Eq. (7), where the concentration of the API in the liquid phase [g-solid API/g-neat solvent] is the fourth state and only output, as in Eq. (8).

|  |  |
| --- | --- |
|  | (7) |
|  | (8) |

The particles are assumed to nucleate at rate [#/(kg-neat solvent.s)] with size [m], volume shape factor [-], and true crystal density [kg/m³], whereas is the neat (anti)solvent stream mixture with mass fraction composition . For surface-integration controlled crystallization, both and B are nonlinear functions of supersaturation (Eq.  10 and 11). Lastly, the solvent mixture volume is estimated by , where , , and are the mass, mass fraction and density of solvent component .

* 1. Case study: OED for parameter estimation of crystallization kinetics

The model is used to obtain the temperature and dosing trajectories that optimize under constraints of similar complexity to an industrial situation as in Eq. (9). This case maximizes the determinant of (i.e., the D-criterion: .

|  |  |  |
| --- | --- | --- |
|  for Subject to:     |  | (9) |

The specified nodes are uniformly spaced in time, where is the node index. The controls were discretized in 10 nodes defining piecewise linear profiles. The initial concentration is optimized along with the control trajectories to leverage data on metastable zone at initial condition. Similarly, the relative seed mass is optimized given the initial moments scale linearly with seed mass. The solution is obtained by single shooting in MATLAB® 2022b. The augmented ODE system is solved through ode23s, and optimization is compiled in FORTRAN (called internally by Intel oneAPI 2022) through a combination of multi-start optimization using the COBYLA (Powell, 1994) and SIMPLEX algorithms. Levenberg-Marquardt, genetic algorithm, particle-swarm etc., were applied but COBYLA was faster and achieved lower minima.

Growth and secondary nucleation are given by Eq. (10), with parameters , , , .

|  |  |
| --- | --- |
|  for , otherwise | (10) |

The supersaturation is computed by Eq. (11), where the denominator refers to the saturated state, and and are the API activity coefficient and molar fraction. The procedure to obtain the activity coefficients is detailed elsewhere (Quilló et al., 2021).

|  |  |
| --- | --- |
|  | (11) |

The API solubility is given by the Van’t Hoff Jouyban-Acree model (Eq. 12), with T as temperature [K], as the mass fraction of solvent and as parameters, as in **Table 1**.

|  |  |
| --- | --- |
|  | (12) |

**Table 1**. Van't Hoff Jouyban-Acree solubility parameters for the case study.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Analytical derivatives were used for the sensitivity equations in the Jacobian to improve accuracy and reduce the computational burden. The value of the concentration variance is approximated by the model mean squared error (assumed to be 8.74 10-7). The sampling interval () for the states is assumed to be 5 min, in an experiment of  = 16h. The initial solvent composition is , and , and pure solvent 2 is dosed. Other information for the case-study is supplied in **Table 2**.

**Table 2.** Summary of model parameters used for OED and process parameters for the case study.

|  |  |  |  |
| --- | --- | --- | --- |
|  = 14.78 |  = 7.18 | = -5.13 | = 2.03 |
| = 1300 kg/m3 |  = 0.1 | = 0.21 kg/kg | = 0.303 kg |
| = 0.1% |  = 1.0% |  = 0.303 kg/kg | = 0.333 kg/kg |
|  = 2.05107 # |  = 2.83 102 m |  = 9.6710-3 m2 |  = 400 mL |
|  = 33 °C |  = 60 °C | = -1 °C/min |  1 |
|  = 1.2 mL/min |  = 1.01 |  |  |

**Figure 1** shows it is D-optimal to produce supersaturation peaks at the start of the experiment (in an attempt) to distinguish growth and birth rate in time. This is followed by a nearly constant supersaturation period and secondary peaks caused by the dosing at the end to satisfy constraints. The sensitivities behave similarly, with a small offset due to the rate of increase of during the simulation. Moreover, (re)estimating and is easier than and during subsequent data collection due to their larger sensitivities.

|  |  |
| --- | --- |
| A graph of a heat solution  Description automatically generated | A graph of a function  Description automatically generated |
| A graph of a mass of added stream  Description automatically generated | A graph with red lines and numbers  Description automatically generated |

**Figure 1.** D-optimal *in silico* experiment, with = 0.314 kg/kg and 0.187% seed load. (top-left) Evolution of the concentration, solubility, and supersaturation, (top-right) output sensitivities, (bottom-left) control policy for temperature and dosing, (bottom-right) growth and nucleation rates.

* 1. Conclusions

While the possibility of more precise parameters estimates is a tempting route for data collection, OED limitations are evidenced by the multitude of combinations of kinetic rates models. OED epitomizes a test for model quality, which is not always satisfied due to poor statistical properties of the overall model structure or the absence of a model output relating to the PSD to partially decorrelate parameters from different phenomena.

The D-efficiency of the obtained design (with )with respect to an analogous experiment with one-step linear cooling and dosing to the values of the final nodes of the D-optimal experiment (feasible, with **)** with is defined as and was computed as 3.71, meaning the analogous experiment would have to be replicated that many times for a similar reduction in parameter CI volume (i.e., assuming the parameters do not change). However, the low values of the determinants ( for the OED) indicate some level of parameter unidentifiability. The same behavior is observed in other mechanistic functions for growth such as Burton-Cabrera-Frank and birth-and-spread, or expressions for nucleation.

OED is a useful addition to the strained R&D timelines, reducing the experimental effort through *in silico* designs. However, the screening experiments need to generate enough data for an imprecise model and comprehensively define a feasibility domain. Moreover, the correlated kinetic expressions used in PBMs, likely mismatch between a true follow-up experiment and the *in silico* experiment, along the experimental hurdles, pose a challenge to model discrimination and identification, prerequisites for successful OED.

* 1. Acknowledgements

This work was supported by VLAIO (Agentschap Innoveren & Ondernemen) and Janssen Pharmaceutica [Baekeland grant number HBC.2020.2214].

References

Á. Orosz, M.H. Bosits, É. Pusztai, H. Pataki, Z. Szalay, Á. Demeter, B. Szilágyi, 2023. Diastereomer salt crystallization: Comprehensive process modeling and DoE-driven comparison of custom-coded and user-friendly simulators. Chem. Eng. J. 473, 145257.

G.L. Quilló, J.F.M. Van Impe, A. Collas, C. Xiouras, S.S. Bhonsale, 2023a. Dynamic Optimization of Active Pharmaceutical Ingredient (Semi-)Batch Crystallization using Population Balance Modelling, in: Computer Aided Chemical Engineering. Elsevier Masson SAS, pp. 1495–1500.

G.L. Quilló, S. Bhonsale, B. Gielen, J.F. Van Impe, Collas, A., Xiouras, C., 2021. Crystal Growth Kinetics of an Industrial Active Pharmaceutical Ingredient: Implications of Different Representations of Supersaturation and Simultaneous Growth Mechanisms. Cryst. Growth Des. 21, pp. 5403–5420.

G. L. Quilló, S. Bhonsale, A. Collas, C. Xiouras, J.F.M. Van Impe, 2023b. Iterative model-based optimal experimental design for mixture-process variable models to predict solubility. Chem. Eng. Res. Des. 189, pp. 768–780.

H. Kilari, Y. Barhate, Y.-S Kang, Z.K. Nagy, 2023. A Systematic Framework for Iterative Model-Based Experimental Design of Batch and Continuous Crystallization Systems, in: Computer Aided Chemical Engineering. pp. 1501–1506.

M.J.D. Powell, 1994. A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation, in: Advances in Optimization and Numerical Analysis. Springer Netherlands, Dordrecht, pp. 51–67.

S. Bhonsale, P. Nimmegeers, S. Akkermans, D. Telen, I. Stamati, F. Logist, J.M.F. Van Impe, 2022. Optimal experiment design for dynamic processes, in: Simulation and Optimization in Process Engineering. Elsevier, pp. 243–271.

Z.K.Nagy, M. Fujiwara, R.D.Braatz, 2008. Modelling and control of combined cooling and antisolvent crystallization processes. J. Process Control 18, pp. 856–864.